Report for Development of a Census Array and Evaluation of the Array to Detect Biothreat Agents and Environmental Samples for DHS

Contributors:
Crystal Jaing
Shea Gardner
Kevin McLoughlin
James Thissen

Lawrence Livermore National Laboratory (LLNL), Livermore, CA

Principal Investigator and Correspondent
Crystal Jaing
925-424-6574, jaing2@llnl.gov

Paul Jackson
(925) 424-2729, jackson80@llnl.gov

SUMMARY

The objective of this project is to provide DHS a comprehensive evaluation of the current genomic technologies including genotyping, Taqman PCR, multiple locus variable tandem repeat analysis (MLVA), microarray and high-throughput DNA sequencing in the analysis of biothreat agents from complex environmental samples. This report focuses on the design, testing and results of samples on the Census Array.

We designed a Census/Detection Array to detect all sequenced viruses (including phage), bacteria (eubacteria), and plasmids. Family-specific probes were selected for all sequenced viral and bacterial complete genomes, segments, and plasmids. Probes were designed to tolerate some sequence variation to enable detection of divergent species with homology to sequenced organisms, and to be unique relative to the human genome. A combination of "detection" probes with high levels of conservation within a family plus "census" probes targeting strain/isolate specific regions enabled detection and taxonomic classification from the level of family down to the strain. The array has wider coverage of bacterial and viral targets based on more recent sequence data and more probes per target than other microbial detection/discovery arrays in the literature. We tested the array with purified bacterial and viral DNA/RNA samples, artificial mixes of known bacterial/viral samples, spiked DNA against complex background including BW aerosol samples and soil samples, and environmental samples to evaluate the array's sensitivity and forensic capability. The data were analyzed using our novel maximum likelihood software. For most of the organisms tested, we have achieved at least species level discrimination.

METHODS

1. Sequence analysis and microarray probe design to develop "census" array

We included two types of probes on the "census" array: detection probes and census probes. *Detection* probes are conserved across multiple sequences from within a family or family-unclassified viral group, but not conserved across families or kingdoms (i.e. they are unique to a family). Such probes aim to detect known organisms or discover novel organisms that have not been sequenced but which possess some sequence homology to organisms that have been sequenced, particularly in those regions found to be conserved among previously sequenced members of that family. We have previously design a Lawrence Livermore Microbial Detection Array using this approach (1). These conserved probes may identify an organism to the level of genus or species, for example, but may lack the specificity to pin the identification down to strain or isolate. *Census* probes, in contrast, represent the least conserved regions, that is, the most strain or isolate specific probes. Such census probes aim to fill the goal of providing higher level discrimination/identification of known species and strains to facilitate forensic resolution, but may fail to detect novel organisms with limited homology to sequenced organisms. We included both types of probes on the census array to maximize the capability to detect both well-characterized and novel microbes and to facilitate high confidence classification at both higher (family) and lower (species and strain) taxonomic levels.

The array design process is diagrammed in Figure 1. We downloaded all sequences, including complete genomes and sequence fragments (genes, noncoding regions, etc.),

organized by family, for all bacteria and viruses, from NCBI Genbank, Integrated Microbial Genomics (IMG) project at the Joint Genome Institute, The Comprehensive Microbial Resource (CMR) at the JC Venter Institute, and The Sanger Institute in the United Kingdom, with some additional proprietary whole-genome data from collaborators. Bacteria were those under the superkingdom Bacteria (eubacteria) taxonomy node at NCBI, and did not include the Archaea. Sequence data for complete genomes, viral segments, and plasmids was current as of August 2009, and for sequence fragments as of January 2009. Table 2 summarizes the number of families, species, genomes, and sequence fragments represented on the array. The process of downloading the sequence data into curated groups required more than a week, with automated scripts running 24x7.

Figure 1: Diagram of the census/detection array design process
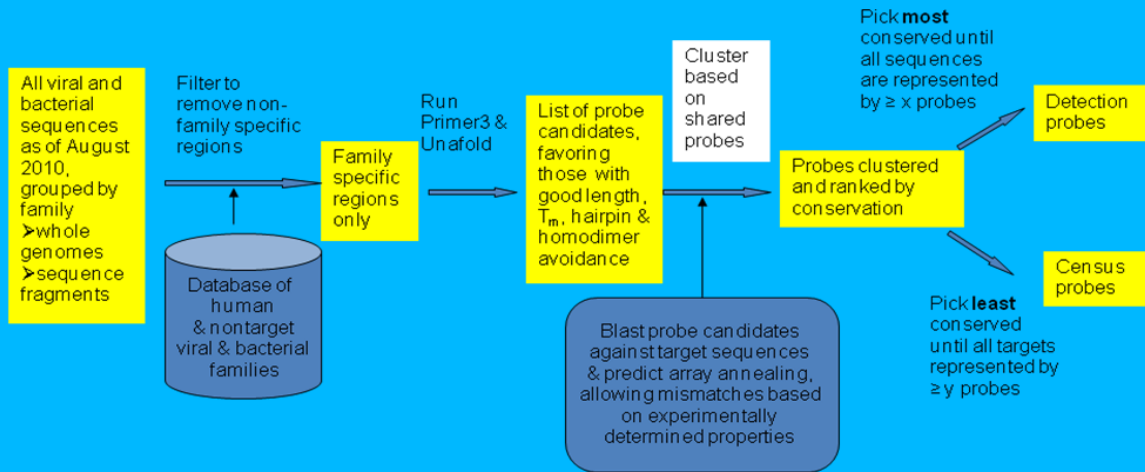
We began the array design process by identifying the family specific regions in each family. In prior work, we have found that the length of longest perfect match (PM) is a strong predictor of hybridization intensity, and that for probes at least 50 nt long, PM≤20 bp have signal less than 20% of that with a perfect match over the entire length of the probe (2). This is similar to results from a systematic study of viral probe hybridization characteristics by (3). Therefore, for each target family we eliminated regions with perfect matches to sequences outside the target family. Using the suffix array software vmatch (4), perfect match subsequences of at least 17 nt long present in non-target viral families or 25 nt long present in the human genome or nontarget bacterial families were eliminated from consideration as possible probe subsequences. Sequence similarity of probes to non-target sequences below this threshold was allowed, but could be accounted for using the statistical algorithm described below.

From these family-specific regions, we designed probes 50-66 bases long for one family at a time using the methods described in (2). Briefly, we generated candidate probes using MIT's Primer3 software (5), followed by T_m and homodimer, hairpin, and probe-target free energy (ΔG) prediction using Unafold. Candidate probes with unsuitable ΔG's or T_m's were excluded as described in (2). Desirable range for these parameters was 50≤length≤66, T_m≥80°C, 25%≤GC%≤75%, $\Delta G_{homodimer}$= ΔG of homodimer formation > 15 kcal/mol, $\Delta G_{hairpin}$=ΔG of hairpin formation > -11 kcal/mol, and $\Delta G_{adjusted}$

$= \Delta G_{complement} - 1.45 \, \Delta G_{hairpin} - 0.33 \, \Delta G_{homodimer} \leq -52$ kcal/mol. An additional minimum sequence complexity constraint was enforced, requiring a trimer frequency entropy of at least 4.5 (calculation described below). If fewer than a minimum number of candidate probes per target sequence passed all the criteria, then those criteria were relaxed to allow a sufficient number of probes per target. To relax the criteria, first candidates that passed the primer3 criteria but failed the Unafold filters were allowed. If no candidates passed the primer3 criteria, then regions passing the target-specificity (e.g. family specific) and minimum length constraints were allowed.

Next, we BLASTed these candidates against the family of target sequences from which they were designed to predict the targets that should be represented by each candidate. A target was considered to be represented if a probe matched it with at least 85% sequence similarity over the total probe length, and a perfectly matching subsequence of at least 29 contiguous bases spanned the central base of the probe (it could be off center, so long as it spanned the middle position). We ranked the probe candidates by their level of conservation, that is, how many targets they were predicted to represent. From here, we followed two contrasting strategies to pick 1) detection probes, and 2) census probes.

For detection probes, we selected probes in *decreasing* order of the number of targets represented by that probe (i.e. probes detecting more targets in the family were chosen preferentially over those that detected fewer targets in the family).For probes that tied in the number of targets represented, a secondary ranking was used to favor probes most dispersed across the target from those probes which had already been selected to represent that target. The probe with the same conservation rank that occurs at the farthest distance from any probe already selected from the target sequence is the next probe to be chosen to represent that target.

For census probes, the process was similar except that we selected probes in *increasing* order of the number of targets represented by that probe (i.e. probes detecting fewer targets in the family were chosen preferentially over those that detected more targets in the family). The same criterion as described above was imposed to maximize positional dispersion of multiple probes across each target sequence. A minimum of 5 probes per target sequence were included. For sequences that diverged from other members of the family or that clustered as a highly conserved subgroup (e.g. multiple sequences from the same outbreak), the detection and census probes could be the same. Duplicate probes were removed in the final chip design.

We included 5-30 detection probes per target and 1-10 census probes per target depending on the array density. Several versions of the census and combined detection/census arrays were designed that differ in density, and thus cost. The standard census array fits on the 388K NimbleGen design (Table 2). The Nimblegen 2.1M and Agilent 1M formats allowed more probes per target sequence. These higher density format arrays were not used for the NGFA-5 sample analysis. Detection probes were designed for all targets, both complete genomes and plasmids and sequence fragments. Census probes were design for all viruses, both complete and partial sequences, and for all complete bacterial genomes and plasmids. Census probes were also designed for sequence fragments for the ~240 bacterial families with less available sequence data (<~150MB), although array density limitations did not allow us to include census probes

for the sequence fragment data for the 32 families with the most available sequence data (~200 MB-2.5 GB and with thousands of sequences), since those families were already so well-represented by the copious detection probes as well as census probes for the numerous complete genomes. Moreover, these partial sequences included many extremely highly conserved rRNA genes which are inappropriate for strain discrimination. Additional probes representing the partial sequences for these already heavily represented families was thought to be unnecessary for the goal of censusing for strain discrimination.

Approximately 1,000- 3,500 random negative control sequences length and GC% matched to the target probes were included. These had no appreciable homology to known sequences based on BLAST similarity, and were used to assess background hybridization intensity. We included probes that tile across the Biowatch amplicons for the viral hemorrhagic fevers, to increase sensitivity in case the array were used to confirm a Biowatch hit for these organisms. Probes were also included to represent unpublished viral sequence fragments provided by our collaborators at the San Francisco Blood Systems Research Institute (abbreviated as SFBB for SF Blood Bank).

Table 1: Summary of sequences represented on Census array.

Number of Targets	*Virus*	*Bacteria*
Families	80	274
Groups w/out Family Classification	48	65
Species with complete genome, plasmid, or segment data	2530	1290
Species with any sequence data, including sequence fragments	5719	14765
Sequences Classified as to Family	171264	728467
Sequences UnClassified as to Family	6996	56251
Complete genomes, segments, or plasmids	55803	4122

Table 2: Types of microarray probes for the Standard 388K census array

# of Probes	Probe Type	Comments
380088	Bacteria and virus census and detection probes	Census probes: 5 pps for all viral sequences and bacterial whole genomes and plasmids, 1 pps for bacterial sequence fragments from 248 families. Detection probes: 5 pps for all sequences (breakdown between census and detection is not relevant due to overlap between these sets)

	Hemorrhagic fever virus	Tiled across amplicons with 50% overlap
1821	Biowatch amplicons	between probes (2x coverage)
1860	SFBB sequences	Tiled across amplicons with 0% overlap between probes (1x coverage)
1235	random controls	
385004	Total	

2. DNA Extraction from pure bacteria and environmental samples

2.1 DNA extraction from *B. anthracis* and *B. thuriengiensis* strains.
Genomic DNA from *B. anthracis* Ames, Sterne and A0382 and *B.thuringiensis* Israrensis HD500, kurstaki ATCC 33679, *Y. pestis, F. tularensis, B. pseudomallei* were obtained from LLNL collections. Genomic DNA from *B. mallei, B.abortus, B. melitensis* and *B. suis* were obtained from NBACC.

2.2 DNA extraction from BioWatch Filters
PSU filters from the NCR Laboratory were received from the BioWatch group at LLNL. One quarter of each filter had previously been excised at NCR, so only ¾ of each filter was available. One week of filters was collected from each season: Spring (4/20-4/26/09), Summer (7/19-7/25/09), Fall (10/25-10/31/09), and Winter (1/22/09-1/28/09). For each day, 7-11 "clean" filters were extracted (49-77 per week). Filters were determined to be "dirty" if they had an abundance of soot and dirt captured on their surface.

The ¾ PSU filters were cut into 5 roughly equal pieces using sterile equipment. Up to 24 filters were placed into a 50mL conical tube. 30mL of 100mM phosphate buffer (pH 7.4), 0.05% (v/v) Tween 80 was added to each 50mL tube. The conical tubes were vortexed for 30 seconds and placed on a rocking shaker for 15 minutes. The 30 second vortexing and 15 min shaking was repeated an additional 3 times for a total of 1 hour of washing. The filters were removed from the tube and remaining solution was centrifuged at 3200 x g for 30 minutes at 5°C. Following centrifugation, the supernatant was removed and discarded.

To complete the DNA purification, components of the UltraClean Soil DNA Isolation Kit #12800 from MoBio (Carlsbad, CA) were utilized. The remaining pellet was resuspended with the following solutions added in this order: 100µL TE buffer, 350µL MoBio Bead Solution, 60µL MoBio Solution S1, and 200µL MoBio Inhibitor Removal Solution. A 2mL screw cap tube was loaded with 500mg each of 106 and 500mm zirconia/silica beads. The entire 700µL of resuspended pellet was added to the 2mL bead tube. The samples were bead-beated at max speed for 2 minutes. Following bead-beating, the tubes were centrifuged at 10,000 x g for 30 seconds. The entire supernatant (~450µL) was transferred to a sterile 2mL tube for further extraction.

To the supernatant, 250µL of MoBio Solution S2 was added, vortexed for 5 seconds, and incubated at 4°C for 5 minutes. Following incubation the samples were centrifuged for 1 minute at 10,000 x g and the supernatant transferred to a clean 2mL tube. 2 volumes (~1.3mL) of MoBio Solution S3 was added to the supernatant and vortexed for 5 seconds. The vortexed solution was added in 700µL aliquots, until the entire sample is processed, to a MoBio spin filter and centrifuged for 1 minute at 10,000 x g and the flow-

through discarded. The spin filter was washed 3 times by adding 300μL MoBio Solution S4, centrifuged for 30 seconds at 10,000 $x g$, and the flow-through discarded.

The spin filter was centrifuged an additional 1 minute at 10,000 $x g$ to dry the filter. The filter was placed in a new 2mL collection tube and 50μL of MoBio Solution S5 was added to the membrane. The sample was centrifuged at 10,000 $x g$ for 30 seconds and the eluted DNA was retained. The multiple elutions for each season were combined into one large volume. Samples were speed-vacced to ~50% of the starting volume in order to increase the DNA concentration. DNA concentration was determined by the Invitrogen Qubit fluorometer (Carlsbad, CA).

2.3 DNA Extraction from Soil
Soil was collected in the downtown areas of both Oakland, CA and San Francisco, CA. Four samples were collected in each city at various sites. Samples were extracted using the MoBio UltraClean Soil DNA Isolation Kit #12800. The manufacturer's Alternative Protocol (For Maximum Yields) was followed for this work. The only deviation from the protocol was to wash twice (Step 15) with Solution S4 instead of just once as the protocol stated.

Following extraction, 1ng of each extracted DNA was used in a Real-Time PCR assay to test for inhibition. All samples showed a high level of inhibition of PCR. Based on this each extracted DNA was re-extracted starting from Step 12 of the MoBio Alternative Protocol. This additional extraction is intended to remove additional humic acid. DNA concentration was determined by the Invitrogen Qubit fluorometer.

2.4 DNA Extraction from EPA gauze wipes
Gauze wipes were obtained from a group at LLNL that conducted tests with the EPA. Wipes were used to wipe dirty surfaces indoors and were then inoculated with 6.3 x 10^7 CFU of *B. thuringiensis* kurstaki spores. Following inoculation, DNA was extracted from the wipes using the Promega Blood Extraction Kit (Madison, WI) according to manufacturer's instructions. After extraction, 5.0 x 10^4 CFU were amplified and labeled following the procedure in Section 3 below.

2.5 *Bacillus anthracis* Ames DNA spiked in environmental samples
B. anthracis Ames DNA was acquired from the select agent laboratory within LLNL. Sterility test was performed to ensure the DNA is sterile before the DNA was transferred to our laboratory. DNA was quantified using the Invitrogen Qubit fluorometer and copy number was determined. Six concentrations of *B. anthracis* Ames DNA were made in 10 fold serial dilutions from 1-100,000 copies. Each concentration was mixed with 100 pg of extracted DNA from the Spring NCR filters for the aerosol spike experiments or 1 ng extracted DNA from the combination of soil from Oakland and San Francisco for the soil spike experiments.

3. Microarray Processing
An overview of the microarray process is shown in Figure 2.

3.1 Whole Genome Amplification and Purification

The environmental BioWatch filter samples, EPA wipe samples and *B. anthracis* Ames spiked samples were amplified using the Qiagen REPLI-g Midi Kit #150043 (Valencia, CA). This kit is intended to provide uniform whole genome amplification using Multiple Displacement Amplification. Each copy number dilution of *B. anthracis* DNA spiked in either 1ng of soil or 100 pg of aerosol DNA was amplified using this kit according to manufacturer's instructions. Samples were allowed to amplify for 16 hours at 30 ^{o}C. Amplified samples were purified using the Qiagen Qiaquick PCR Purification Columns #28106 according to manufacturer's instructions. Samples were eluted in 40µL of Buffer EB from the Qiagen kit.

Environmental Sample → Isolate DNA/RNA Amplification if needed → Label with fluorescent dye

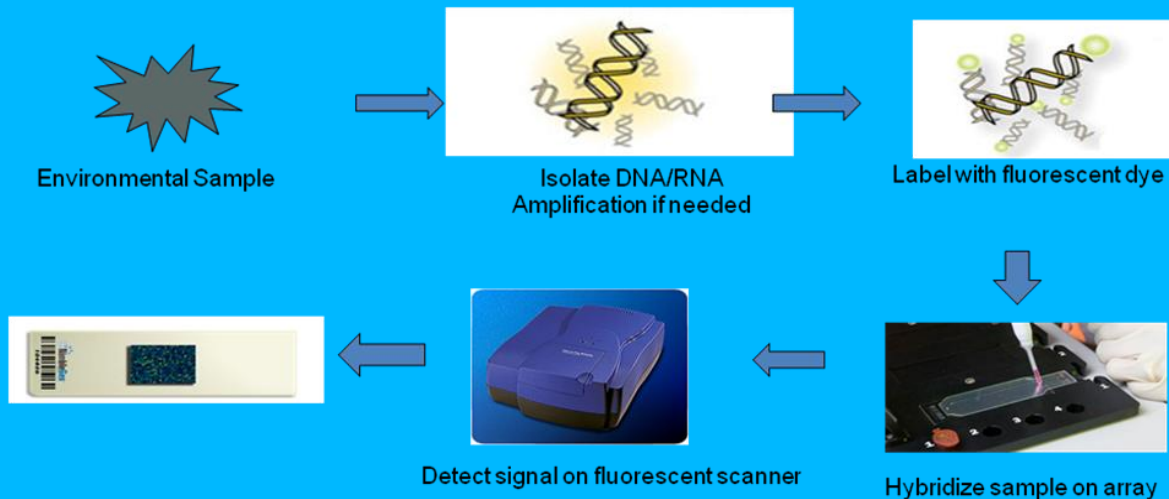Detect signal on fluorescent scanner ← Hybridize sample on array

Figure 2. Microarray hybridization process

3.2 Microarray hybridization
The entire 40µL of amplified product was fluorescently labeled using the Roche NimbleGen One-Color DNA Labeling Kit #05223555001 (Madison, WI) according to the recommended protocols. The DNA was purified after labeling, and hybridized using the NimbleGen Hybridization Kit (Cat. 05583683001) to the LLNL Census Array according to manufacturers' instructions. The microarrays were allowed to hybridize for 17 hours and washed using the NimbleGen Wash Buffer Kit #05584507001 according to manufacturer's instructions. Microarrays were scanned on an Axon GenePix 4000B 5 µm scanner from Molecular Devices (Sunnyvale, CA). The scanned tif image files were aligned using the NimbleScan Version 2.4 software and pair text files were exported for data analysis.

4. Microarray data Analysis
A maximum likelihood analysis method was used to analyze the microbial hits from samples hybridized to the array. The method was recently published in (1). An example of the analysis results is shown in Figure 3 where *B. thuringiensis* kurstaki was run on the Census Array. The right-hand column of bar graphs shows the unconditional and conditional log-odds ratios for each target genome listed at right. The unconditional log-odds is the larger of the two scores; thus the lighter and darker-colored portions represent the unconditional and conditional scores respectively. Targets are color-coded and grouped by taxonomic family, according to the legend at bottom; they are listed within families in decreasing order of conditional log-odds ratio scores. Targets predicted as

likely to be present are indicated in red text. The vertical orange dashed line marks 0 on the log-odds ratio scale.

The left-hand column of bar graphs shows the expectation (mean) values of the numbers of probes expected to be present given the presence of the corresponding target genome. The larger "expected" score is obtained by summing the conditional detection probabilities for all probes; the smaller "detected" score is derived by limiting this sum to probes that were actually detected. Because probes often cross-hybridize to multiple related genome sequences, the numbers of "expected" and "detected" probes often greatly exceed the number of probes that were actually designed for a given target organism.
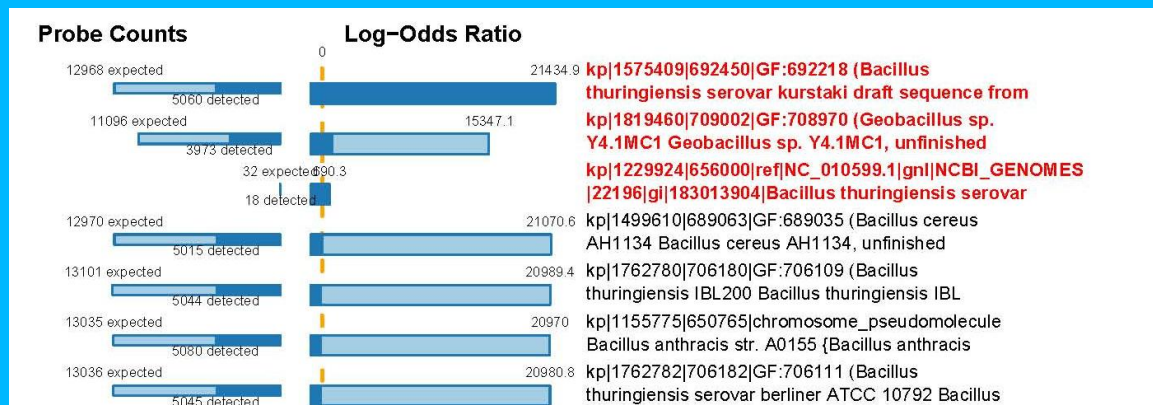

Figure 3. Census Chip result for *B.thurigiensis* kurstaki

RESULTS

1. Hybridization of *B. anthracis* and *B. thuriengiensis* DNAs on the Census Array. The primary goal of array analysis was to identify, for each sample, the organism(s) with known genomic sequence that best explains the pattern of bright (detected) and dark (undetected) probes on the array. Some of the organisms we tested have not been sequenced; for these, our measure of success is whether the analysis identifies the correct species (when other strains of the same species have been sequenced). We isolated genomic DNAs from *B. anthracis* Ames, Sterne and A0382 and DNAs from *B.thuringiensis* Israrensis HD500, kurstaki ATCC 33679. The DNAs were hybridized to the Census Array. The data is shown in Table 3. The Census Array corrected identified each of the DNA to at least the species level. For *B. thuringiensis*, the identification was at the strain level. *B. anthracis* strains were not correctly detected to the strain level. Our analysis algorithm takes into account the combined scores of probe hits to targets. When there is a genomic sequence glued from both draft genomes and plasmids such as the *B. anthracis* A0155, the algorithm assigns a higher score due to the longer length of the sequence. This issue is being addressed by our team so the probe hit scores can be assigned based on the length of the targets, instead of giving higher scores to longer sequences.

Table 3. Census Array results from *B. anthracis*, *B. thuringiensis* and *F. tularensis* DNAs

DNA hybridized on array	Census Array top hit
B. anthracis Ames	*B. anthracis* A0155, *B. anthracis* Sterne
B. anthracis Sterne	*B. anthracis* A0155
B. anthracis A0382	*B. anthracis* A0155
B. thuringiensis Israelensis HD500	*B. thuringiensis* IBL4222
B. thuringiensis Kurstaki ATCC 33679	*B. thuringiensis* kurstaki
F. tularensis holarctica LVS	*F. tularensis* LVS

2. Evaluation of mixtures of biothreat bacterial samples

Genomic DNAs from *B. anthracis* Ames, *Y. pestis* CO92, *F. tularensis* LVS, *Brucella abortis*, *B. pseudomallei* PHLS9 and *B. mallei* 23344 were mixed together in one single sample and hybridized on the 388K census array. The results are shown in Table 4.

All six bacterial species were correctly identified except for *B. mallei* which was not detected. It is not clear why this species was not detected in this experiment. *F. tularensis* LVS was correctly identified at the strain level.

Table 4. Census Array results from a mixture of bacterial DNAs

Actual Species	Actual Strain	Predicted Species	Predicted strain
B. anthracis	Ames	*B. anthracis*	A0193
Y. pestis	CO92	*Y. pestis*	PEXU2
F. tularensis	holarctica LVS	*F. tularensis*	LVS
B.pseudomallei	PHLS9	*B.pseudomallei*	E208
B. mallei	ATCC 23344	*ND*	*ND*
Brucella	abortis	*Brucella*	melitensis

3. Determination of the limit of detection of the Census Array using *B. anthracis* Ames spiked into BioWatch aerosol samples.

We performed limit of detection testing of the Census Array using serially diluted *B. anthracis* Ames spiked into BioWatch aerosol filter extracts that have been subjected to whole genome amplification. Duplicate experiments were run to ensure repeatability and data consistency. 1, 10, 100, 1,000, 10,000, and 100,000 copies of *B. anthracis* Ames were tested. Table 5 below shows results of the census array at each of the *B. anthracis* DNA concentration. When 10 copies of *B. anthracis* DNA were spiked into aerosol sample, only one of the two replicate experiments detected *B. anthracis*, suggesting that there were not enough probes specific to *B. anthracis* detected at this concentration. This experiment suggested that our detection limit for *B. anthracis* Ames could be in the range between 10-100 copies when the DNA was spiked into 100 pg of aerosol DNA sample.

Table 5. Limit of detection of *B. anthracis* Ames DNA spiked in Biowatch aerosol samples.

Amount aerosol filter DNA	100 pg	100 pg	100 pg	100 pg	100 pg	100 pg
Amount *B. anthracis* DNA	560 pg	56 pg	5.6 pg	560 fg	56 fg	5.6 fg
B. anthracis DNA Copy #	100,000 copies	10,000 copies	1000 copies	100 copies	10 copies	1 copy
% BA DNA in aerosol DNA	98.2%	35.9%	5.3%	0.56%	0.06%	0.006%
Census Array top hit	*B. anthracis*	*B. anthracis*	*B. anthracis*	*B. anthracis*	*B. anthracis*	Not detected

4. Determination of the limit of detection of the Census Array using *B. anthracis* Ames spiked into soil samples.

We performed as similar limit of detection testing of the Census Array using serially diluted *B. anthracis* Ames spiked into soil extracts that have been subjected to whole genome amplification. The soils were collected locally in San Francisco and Oakland. Duplicate experiments were run to ensure repeatability and data consistency. 1, 10, 100, 1,000, 10,000, and 100,000 copies of *B. anthracis* Ames were tested. Table 6 below shows results of the census array at each of the *B. anthracis* DNA concentration. When 100 copies of *B. anthracis* DNA were spiked into soil sample, only one of the two replicate experiments detected *B. anthracis*, suggesting that there were not enough probes specific to *B. anthracis* detected at this concentration. This experiment suggested that our detection limit for *B. anthracis* Ames could be in the range between 100-1000 copies when the DNA was spiked into 1 ng of soil DNA sample.

Table 6. Limit of detection of *B. anthracis* Ames DNA spiked in soil samples.

Amount soil DNA	1 ng	1 ng	1 ng	1 ng	1 ng	1 ng
Amount *B. anthracis* DNA	560 pg	56 pg	5.6 pg	560 fg	56 fg	5.6 fg
B. anthracis DNA Copy #	100,000 copies	10,000 copies	1000 copies	100 copies	10 copies	1 copy
% BA DNA in soil DNA	35.9%	5.3%	0.56%	0.06%	0.006%	0.0006%
Census Array top hit	*B. anthracis*	*B. anthracis*	*B. anthracis*	*B. anthracis*	Not detected	Not detected

5. Analysis of DNA samples from previous DHS or EPA exercises to release and detect *B. thuringiensis* spores.

Aerosol filters were collected during the days when *B. thuringiensis* kurstaki was sprayed to control Gypsy Moth in a DHS surrogate study. Gauge wipes were used to wipe dirty indoor surfaces and inoculated with *B. thuringiensis* kurstaki spores in an exercise conducted with EPA. Genomics DNAs were extracted from filters or wipe samples and

run on the Census Array. The results are shown in Table 7. We were able to positively identify *B. thuringiensis* kurstaki using the Census Array from both the air filter samples and the gauge wipe samples.

Table 7. Detection of *B. thuringiensis* from environmental air or wipe samples.

Sample	Air filters collected around Gypsy Moth control study	Gauge wipes collected during an EPA exercise
B. thuringensis kurstaki specific Taqman asay	Average Ct = 22.77 ± 0.27	Average Ct = 30.70 ± 0.31
Census array top hits	*B. thuringiensis* kurstaki *Bacillus megaterium QM B1551 Magnetospirillum Magnetotacticum MS−1 Thioalkalivibrio sp. Ralstonia pickettii 12D*	*B. thuringiensis* kurstaki *Magnetospirillum magnetotacticum MS−1 Alkalilimnicola ehrlichei Tolumonas auensis DSM*

REFERENCES

1. Gardner, S., Jaing, C., McLoughlin, K. and Slezak, T. (2010) A Microbial Detection Array (MDA) for viral and bacterial detection. BMC Genomics, 11, 668doi:610.1186/1471-2164-1111-1668.

2. Jaing, C., Gardner, S., McLoughlin, K., Mulakken, N., Alegria-Hartman, M., Banda, P., Williams, P., Gu, P., Wagner, M., Manohar, C. et al. (2008) A Functional Gene Array for Detection of Bacterial Virulence Elements. PLoS ONE, 3, e2163.

3. Wong, C., Heng, C., Wan Yee, L., Soh, S., Kartasasmita, C., Simoes, E., Hibberd, M., Sung, W.-K. and Miller, L. (2007) Optimization and clinical validation of a pathogen detection microarray. Genome Biology, 8, R93.

4. Giegerich, R., Kurtz, S. and Stoye, J. (2003) Efficient implementation of lazy suffix trees. Software-Practice and Experience, 33, 1035-1049.

5. Rozen, S. and Skaletsky, H. (2000) In Krawetz, S. and Misener, S. (eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology Humana Press, Totowa, NJ, pp. 365-386.